

What's Spoppin'?

Predicting Popularity of Spotify Tracks
By Erin Vu



Overview



1. Business Problem
2. Data
3. Modeling Process
4. Model Results
5. Conclusion

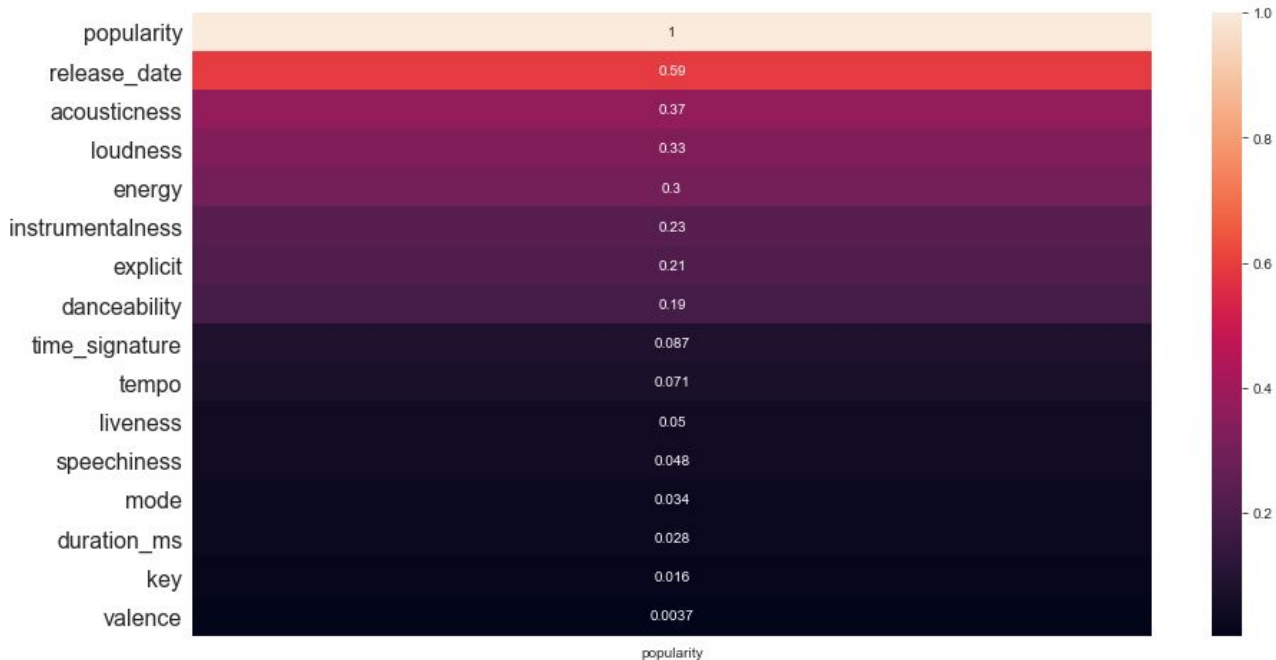
Quality Over Quantity

- Stakeholder: Music production companies and producers
- Dominate streaming/radio charts
- Creating popular songs
- Understanding popular song features
- Using best features to craft songs



Data

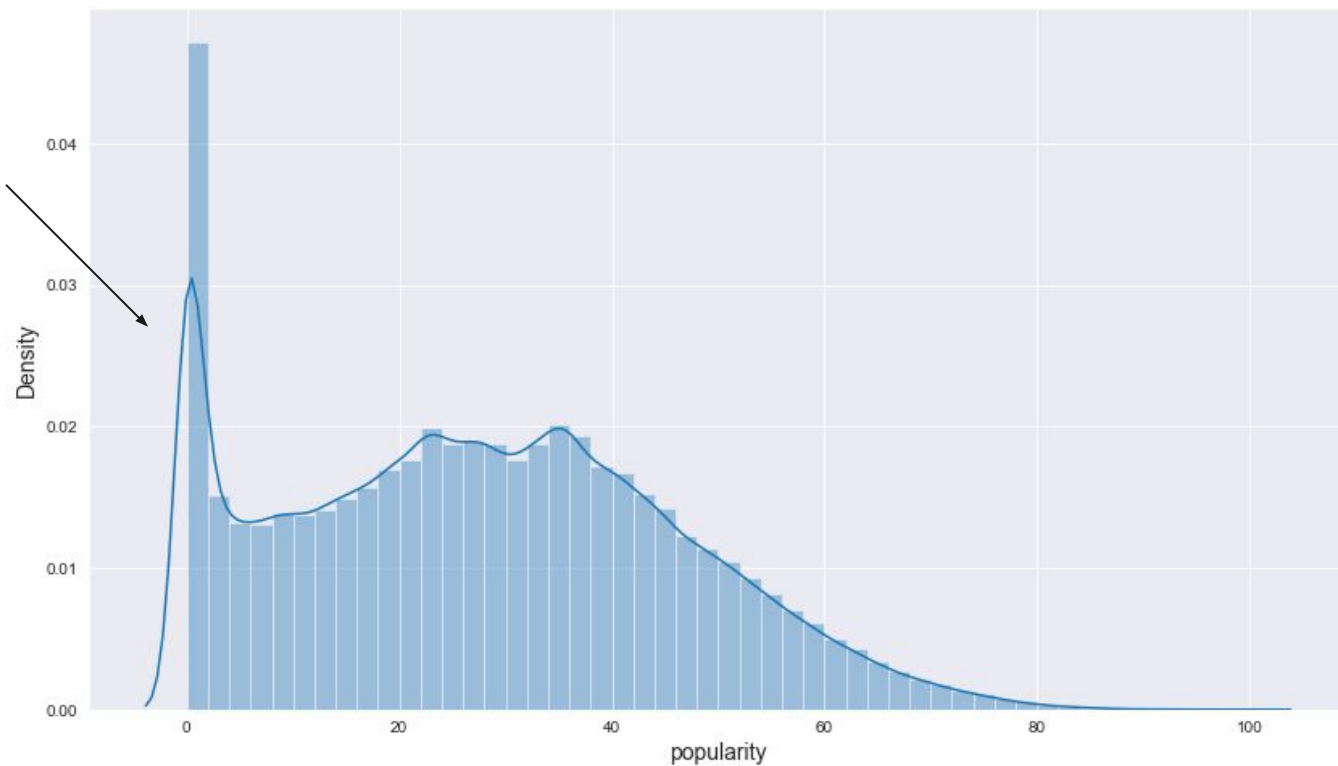
- Spotify Kaggle Dataset
- 1922-2021
- 600k songs
- 20 Features



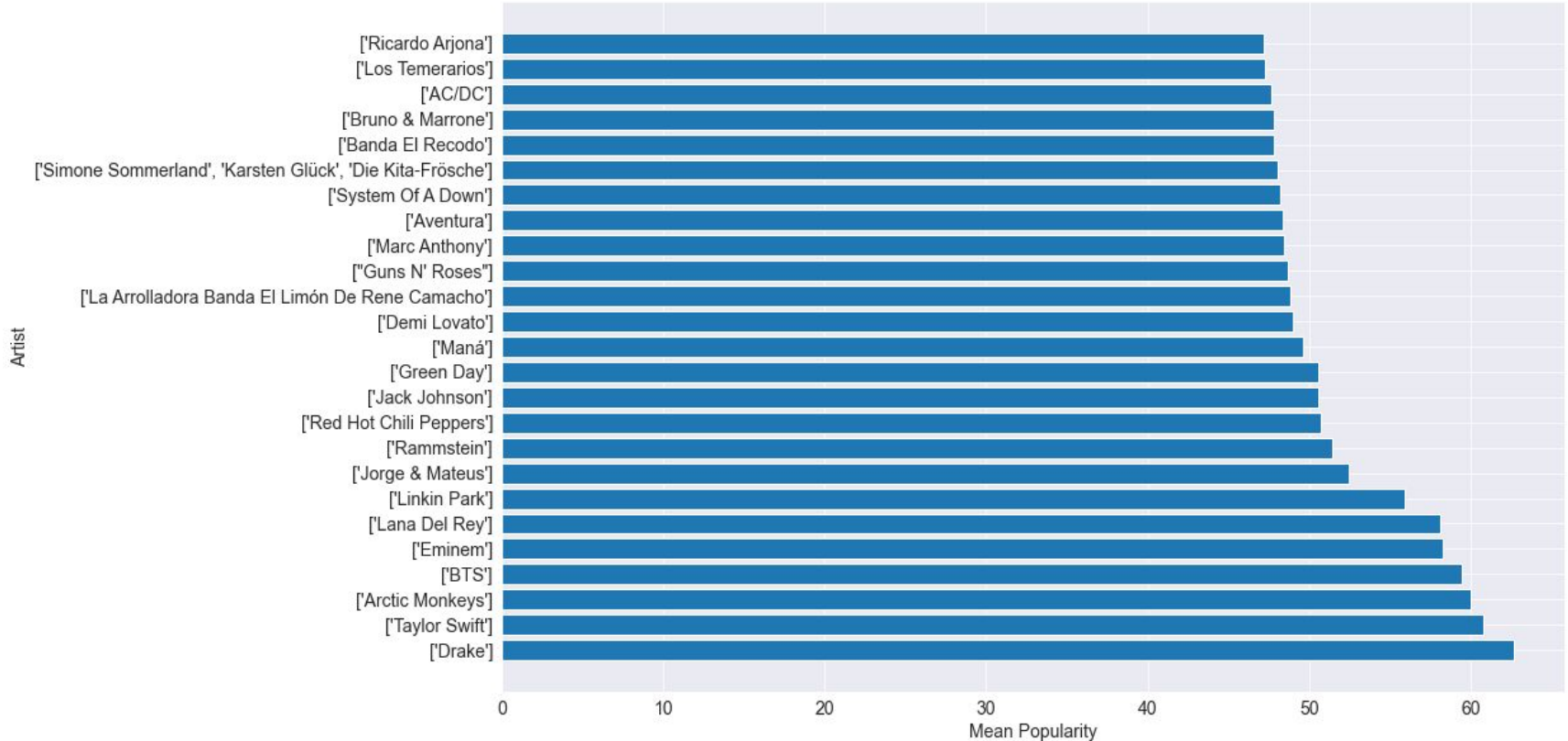
Target: Popularity



43k/~7%
songs with 0
popularity



Top 25 Artists with > 100 Songs

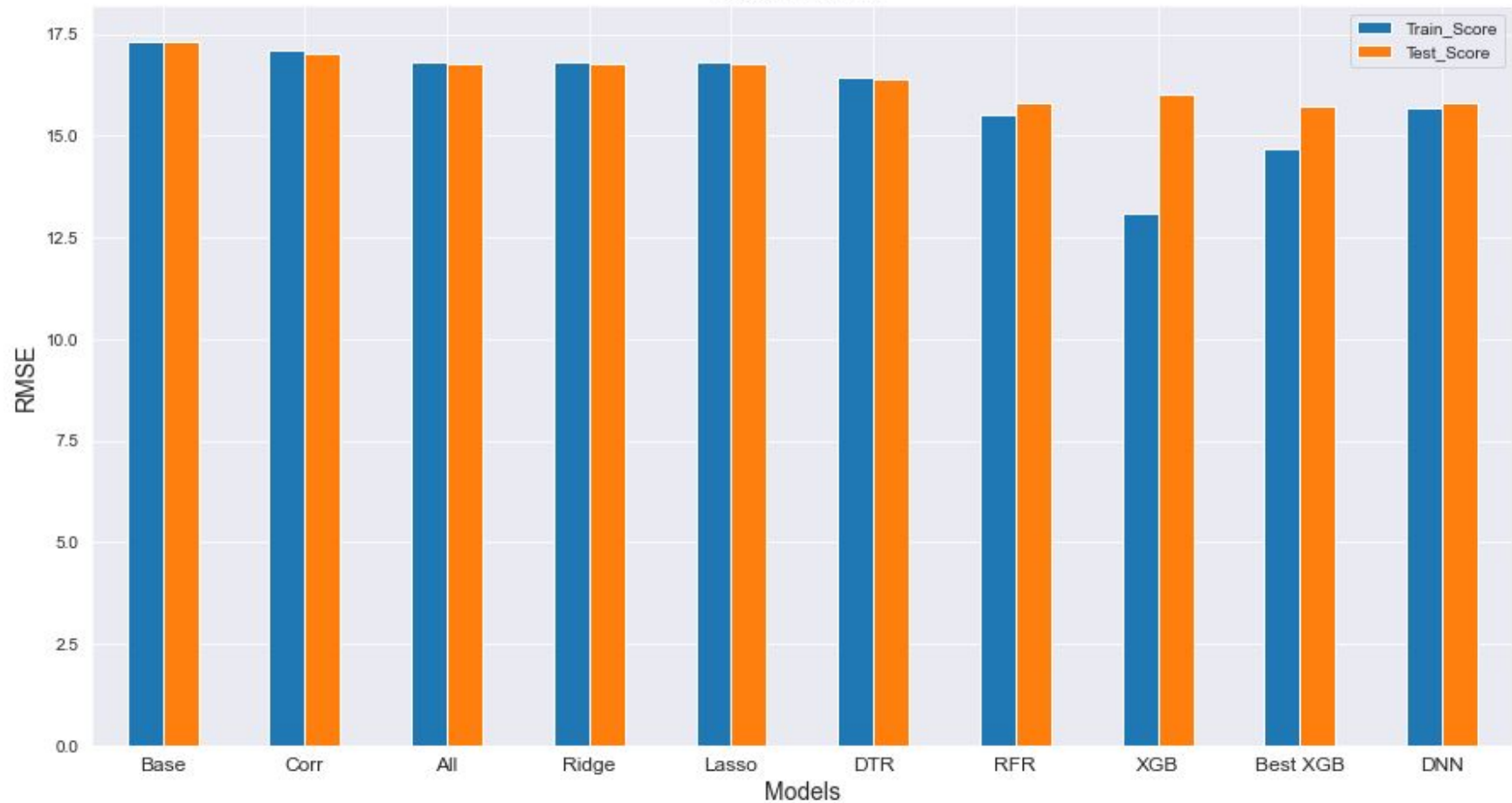


Models

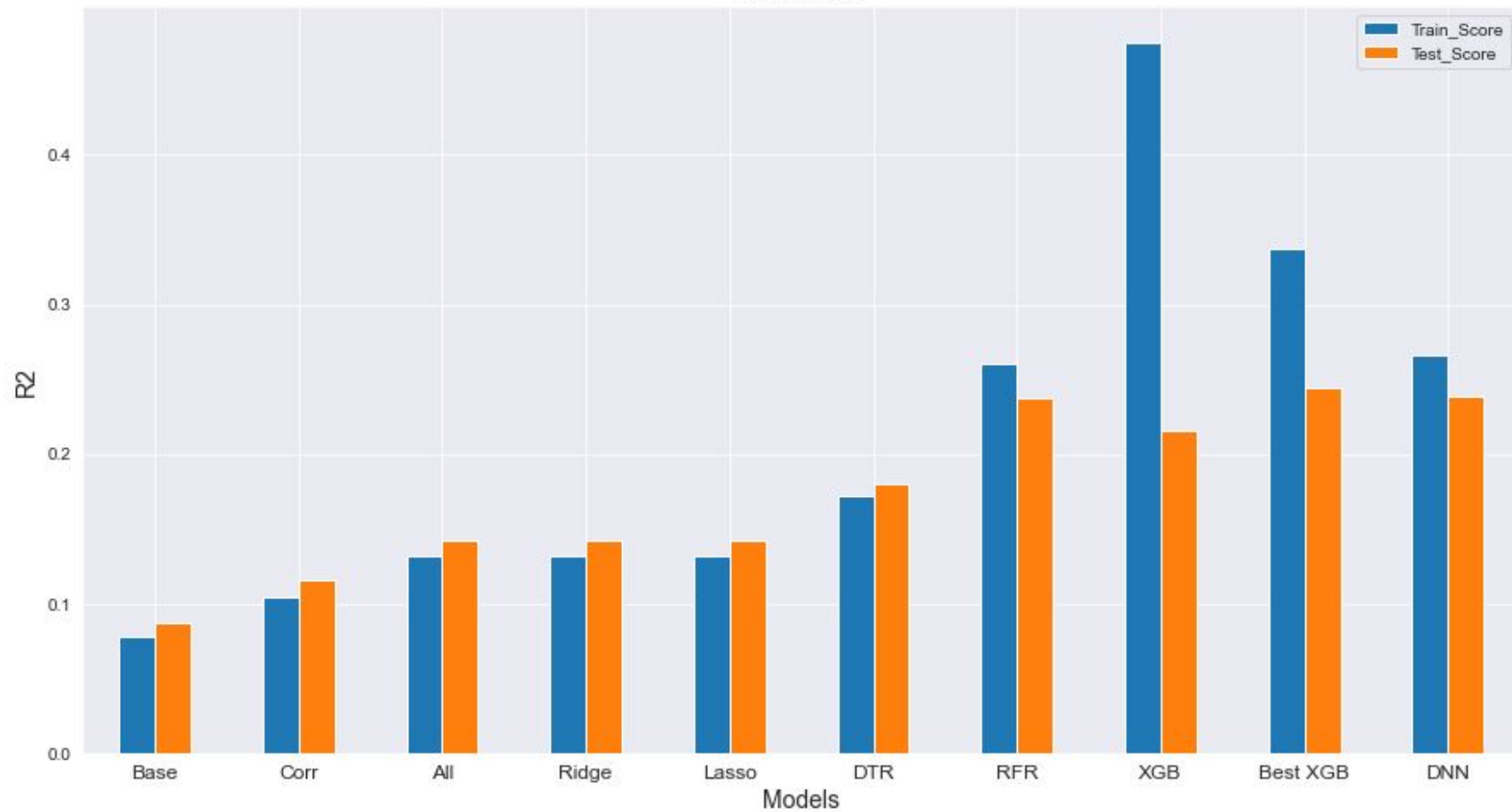


Metric: RMSE
(root mean squared error)

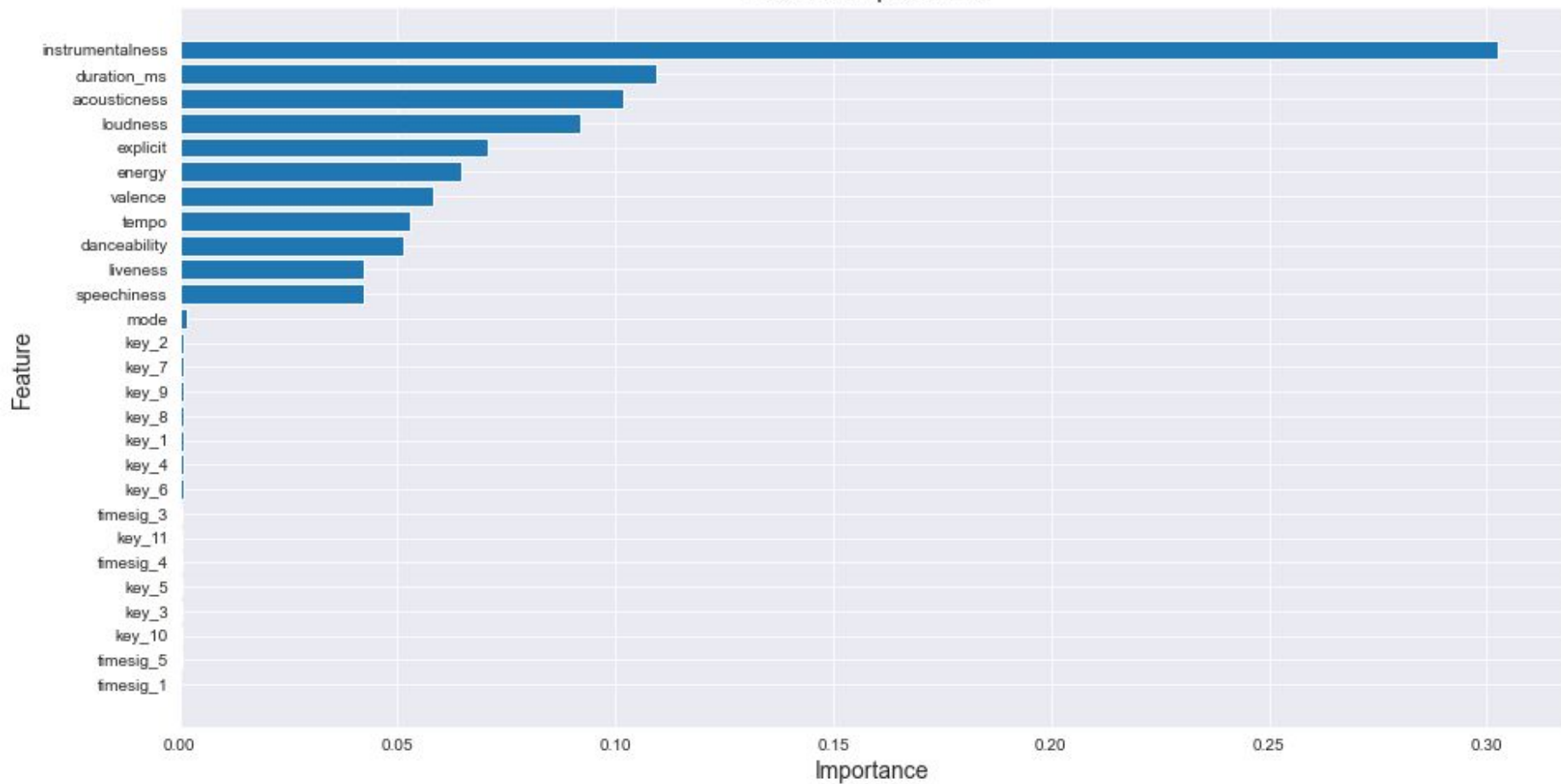
RMSE Scores



R2 Scores



Feature Importances





Conclusion and Next Steps

- Best Model: Random Forest Regression
- Have a good mix of vocals and instruments
- Continue optimizing
- Different types of music
 - Model to predict within genres
 - Subset different time periods

Thank You!

Erin Vu

<https://github.com/ekvu>

Erin.vu94@gmail.com

<https://linkedin.com/in/erin-vu>

